Adobe Acrobat 5.0

# Create searchable PDF document archives

With Adobe® Acrobat® 5.0, it's easy to make a collection of documents available to your entire organization—whether you need to share forms, Web sites, technical documents, or even archived e-mail messages. Save the documents as Adobe PDF files, and then use the Catalog feature in Acrobat to create a searchable full-text index. Distribute the collection on a CD-ROM, or make it available on a server: Using the Search command, you can search the entire collection for specific words or for document information such as keywords, titles, or dates. (In Mac OS X, Catalog and Search are available only in Classic mode.)

1. Prepare PDF documents for indexing.

Save documents in PDF format using the Web capture feature (for Web pages), Acrobat Distiller (for other electronic documents), or the Paper Capture plug-in (for paper documents). Name the documents with eight characters and a .pdf extension if searches will take place from more than one platform (for example, Windows and Mac OS). If you use Windows, make sure that the PDF documents are all on the same hard drive or network volume; on Mac OS, the PDF documents can be on different volumes.

To minimize the need to rebuild the index, ensure that each document is complete, with the appropriate bookmarks, links, and form fields. Searchers have more options if you add a descriptive title, author information, subject, and keywords in the Document Information dialog box for each document.
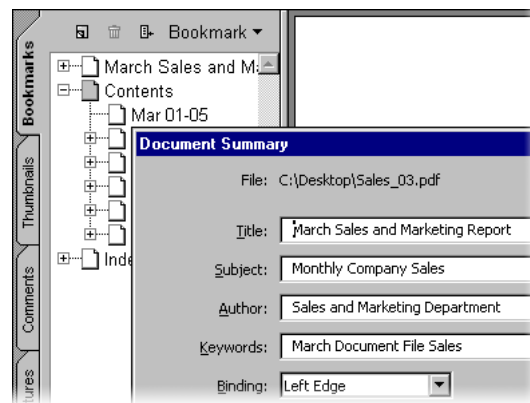
2. Create the index.

In Acrobat, choose Tools > Catalog, and click New Index. Type a title and description for the index. Click Add in the Include These Directories section, and double-click the folders that contain the files you want to index. To exclude subfolders, click Add in the Exclude These Subdirectories section, and select the folders you want to exclude (you can't exclude individual files). Click Options, and select the search options you want to provide. Click Build. Name the index file, and click Save. Catalog reports its progress as it builds the index; click Stop to interrupt the process.

You can use the Options dialog box to exclude specific words (called stopwords) or numbers from the index. Select Optimize For CD-ROM to arrange index files for the fastest access possible on a CD-ROM. Under Word Options, deselect search options that you don't want to provide: Case Sensitive lets the searcher limit matches by capitalization; Sounds Like expands searches for proper names; and Word Stemming finds words that share a stem with the search term.
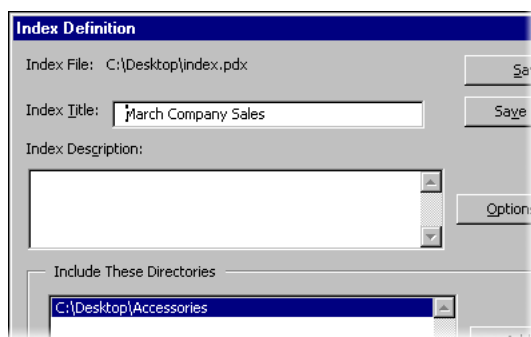
### 3. Move the document collection and index, if desired.

To move the indexed document collection to a CD-ROM, network drive, or intranet, copy or move the folder that contains the document collection, index definition file (PDX), and its support folder.

An index definition contains relative paths between the index definition file and the folders containing the indexed documents. As long as these relative paths remain unchanged, you don't need to rebuild the index after moving the collection.

4. Search the index.

Click the search button on the toolbar. In the Adobe Acrobat Search dialog box, click Indexes, and select each index you want to search. (To search an index that isn't listed, click Add and navigate to it.) Then, type a word or phrase in the Find Results Containing Text box; use *AND* or *OR* to search for multiple words. You can also use wildcard characters: An asterisk (*) represents any number of characters, and a question mark (?) represents a single character. Select options to broaden or limit your search, and then click Search.

Options in the Search dialog box determine which words Acrobat identifies in the search. If the indexes you search were built with these options, you can select Word Stemming to find words that contain the stem of the search word and Sounds Like to find alternative spellings for proper names. Thesaurus finds words that are similar to the search word. Match Case limits the search results to words capitalized like the search word. Proximity limits the results of AND searches to the pair of words that is closest together in each document.
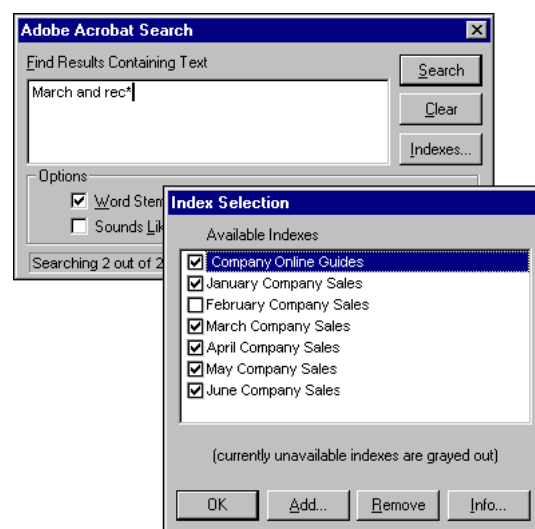
5. View the search results.

In the Search Results dialog box, select a document, and click Info to see its title, subject, and other information. Double-click a document to view it. The search words are highlighted. Click the Next Highlight or Previous Highlight button to view other matches in the document. (Click the arrow next to the Search button to see the Next Highlight and Previous Highlight buttons.)

The Search Results dialog box lists the documents in the index that contain the words you searched for and reports how many documents were searched. The documents are ranked according to the number of search words found in each document. Documents displayed with a solid circle contain the most occurrences of search words; empty circles indicate the fewest occurrences.
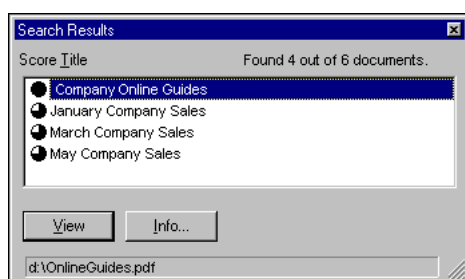
6. Modify your search.

If your search returns too many or no documents, modify your search query. To refine your search, show its results and choose Edit > Search > Query to open the Search dialog box again. Edit the query; then press Ctrl (Windows) or Option (Mac OS)—the button changes from Search to Refine—and click Refine.

To include document information such as titles, keywords, or creation dates in your query, choose Edit > Preferences > General (Windows) or Edit > Preferences (Mac OS), click Search, and select Document Information and Date Filtering under Include in Query. For more information about advanced searching techniques, see the Acrobat 5.0 online Help.
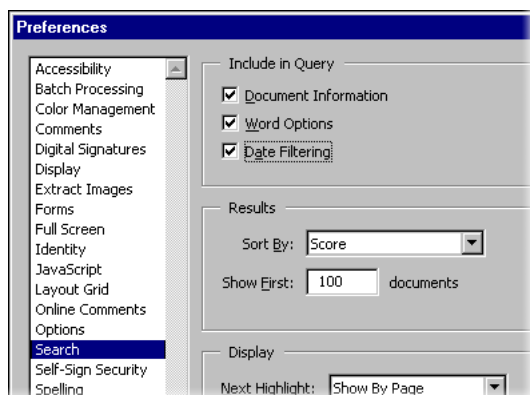
7. Rebuild the index if you change the documents.

Rebuild the index if you remove a document or modify indexed documents: Choose Tools > Catalog, and click Open Index. Select the index definition file (PDX) for the index, and click Build. If you significantly modify the documents, change the stopwords list, or change which search features are supported, purge the index before rebuilding it: Choose Tools > Catalog, and click Open Index. Select the index definition file, and click Purge. After the index is purged, select the index definition file again, and click Build.

When you rebuild an index, Catalog retains entries for deleted documents and for the original versions of changed documents but marks those entries invalid. These additional entries increase the time required for searching, and the disk space the index requires. Purging and rebuilding a changed index can make searching much more efficient. Queries in progress when you purge the index are completed, but new queries return an "Index unavailable for searching" message until the index is rebuilt.